



Facilitate knowledge sharing and create a common understanding about data quality and criteria as well as bringing such quality frameworks into practice

NAPCORE Mobility Data Days

09th November 2023

Dr. Evangelos Mitsakis, Chrysostomos Mylonas, Maria Stavara

Data quality – Legislative basis

EC Delegated Regulation 886/2013 (SRTI)

Article 7

Availability, exchange and reuse of data

4. Public and private road operators and service providers shall ensure the timely renewal and quality of data made available through their access point.

Article 10

Follow-up

2. At the latest 12 months following the entry into force of this Regulation and every calendar year thereafter, Member States shall communicate to the Commission the following information:

(a) the progress they have made in implementing the information service, including the criteria used to define its level of quality and the means used to monitor its quality;

EC Delegated Regulation 2022/492 (RTTI new)

2. The data referred to in paragraph 1 and the corresponding metadata including information on the quality thereof shall be accessible for exchange and re-use by any data user within the Union:

(b) following minimum quality requirements that Member States shall agree upon in cooperation with relevant stakeholders;

5.4. Flexibility for further development of standards and data quality requirements

This same approach has been taken when defining data quality requirements, which need to be developed by cooperating Member States and relevant stakeholders. Existing work on quality requirements and metadata specifications needs to be continuously developed and harmonised.

(21) Member States and ITS stakeholders should be encouraged to cooperate to agree on common definitions of data quality with a view to use common data quality indicators throughout the traffic data value chain, such as the completeness, accuracy and up-to-dateness of the data, the acquisition method and location referencing method used, as well as quality checks applied. They should also be encouraged to work further to establish associated methods of quality measurement and monitoring of the different data types. Member States should be encouraged to share with each other their knowledge, experience and best practices in this field in the on-going and future coordination projects.

Data quality – Prior work

- Focus on priority actions of the EU ITS Directive
- Developed in EIP+ and EU EIP projects

EU EIP
European ITS Platform

Quality of Safety-Related and Real-Time Traffic Information Services

Quality package

EU EIP SA 4.1: Determining Quality of European ITS Services

Version: 2.0
Date: May 15, 2019

Co-financed by the European Union
Connecting Europe Facility
<http://www.its-platform.eu>

EU EIP
European ITS Platform

Multimodal Travel Information Services (MMTIS)

Quality Package

EU EIP 4.1 Task 2: Determining Quality of Multimodal Travel Information Services (MMTIS)

Version 2.0
31 October 2019

Co-financed by the European Union
Connecting Europe Facility
<http://www.its-platform.eu>

EU EIP
European ITS Platform

Quality of Intelligent Truck Parking Services (ITPS)

Quality package

EU EIP SA 4.1: Determining Quality of European ITS Services

Version 2.3
15 October 2019

Co-financed by the European Union
Connecting Europe Facility
<http://www.its-platform.eu>

EU EIP
European ITS Platform

C-ITS

Quality package

EU EIP SA 4.1: Determining Quality of European ITS Services

Version: 1.0
Date: January 21, 2022

Co-financed by the European Union
Connecting Europe Facility
<http://www.its-platform.eu>



Figure: Peter Lubrich

Data quality – NAPCORE WG3 focus

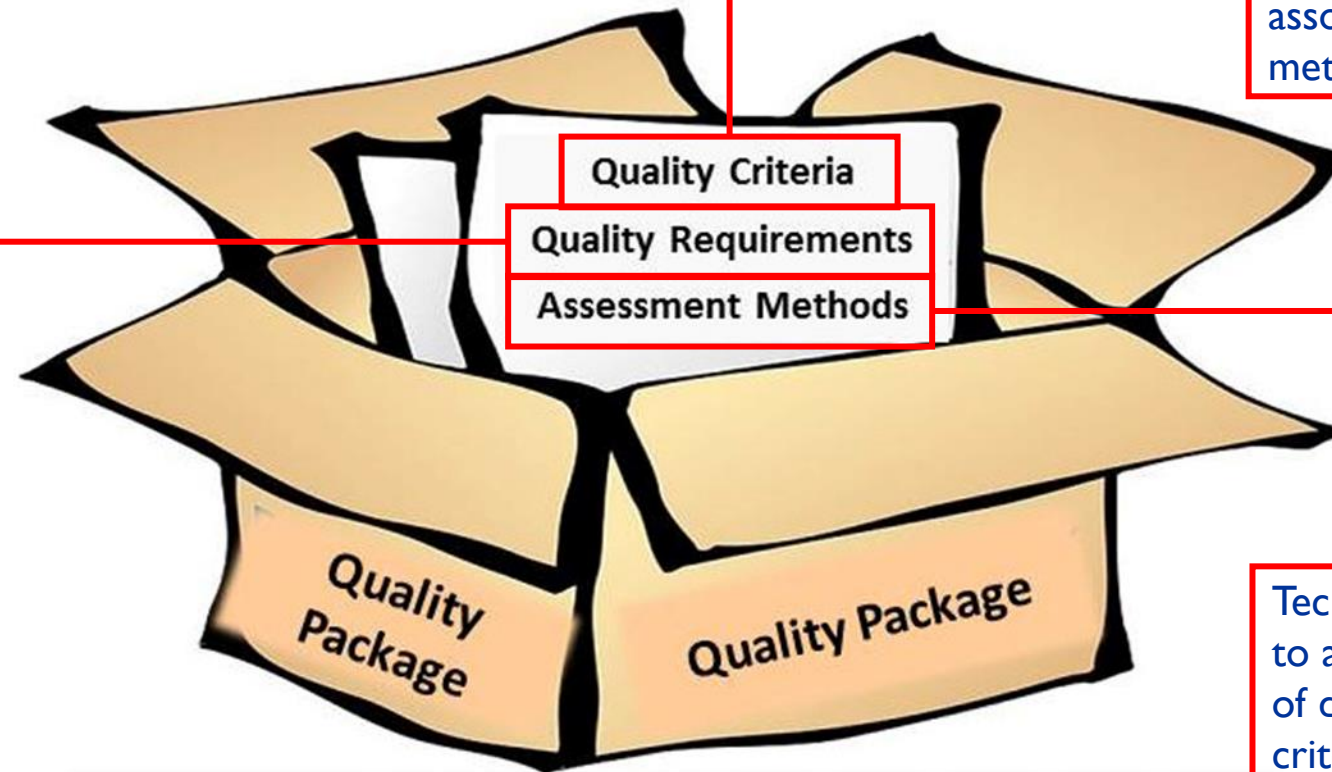
- Extension of Quality Frameworks
 - Development of new QFs focusing on functional domains that are either new or have not been exhaustively covered.
 - Concretization of existing QFs.
- Application & testing on NAP datasets (case studies)

Name of Quality Framework	Scope	Leader
On-street parking data	New framework	DE/BASSt
Alternative fuel data	New framework	PT/Armis
MMTIS data	Concretization of existing	ITxPT
Floating Car Data (FCD)	New framework	GR/CERTH
UVAR data	New framework	PT/Armis
Network Topology Data	New framework	ERTICO
Cross-domain / formal / technical	New framework	CZ/TamTam

Data quality – What is a quality framework?

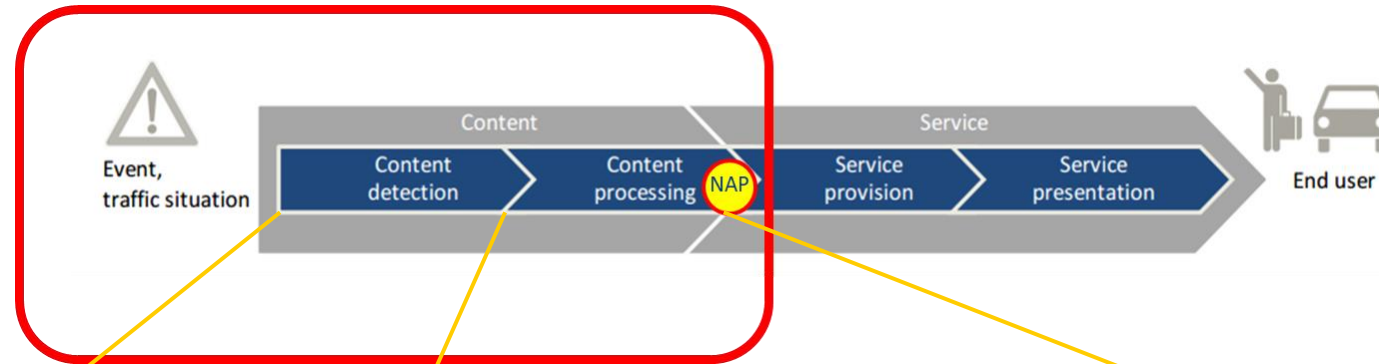
Specific expectations of data quality established by stakeholders and use case needs (criterion-specific).

Specific/measurable aspects that enable data quality assessment. They may be grouped under various quality dimensions (overarching categories) and associated with various quality metrics.



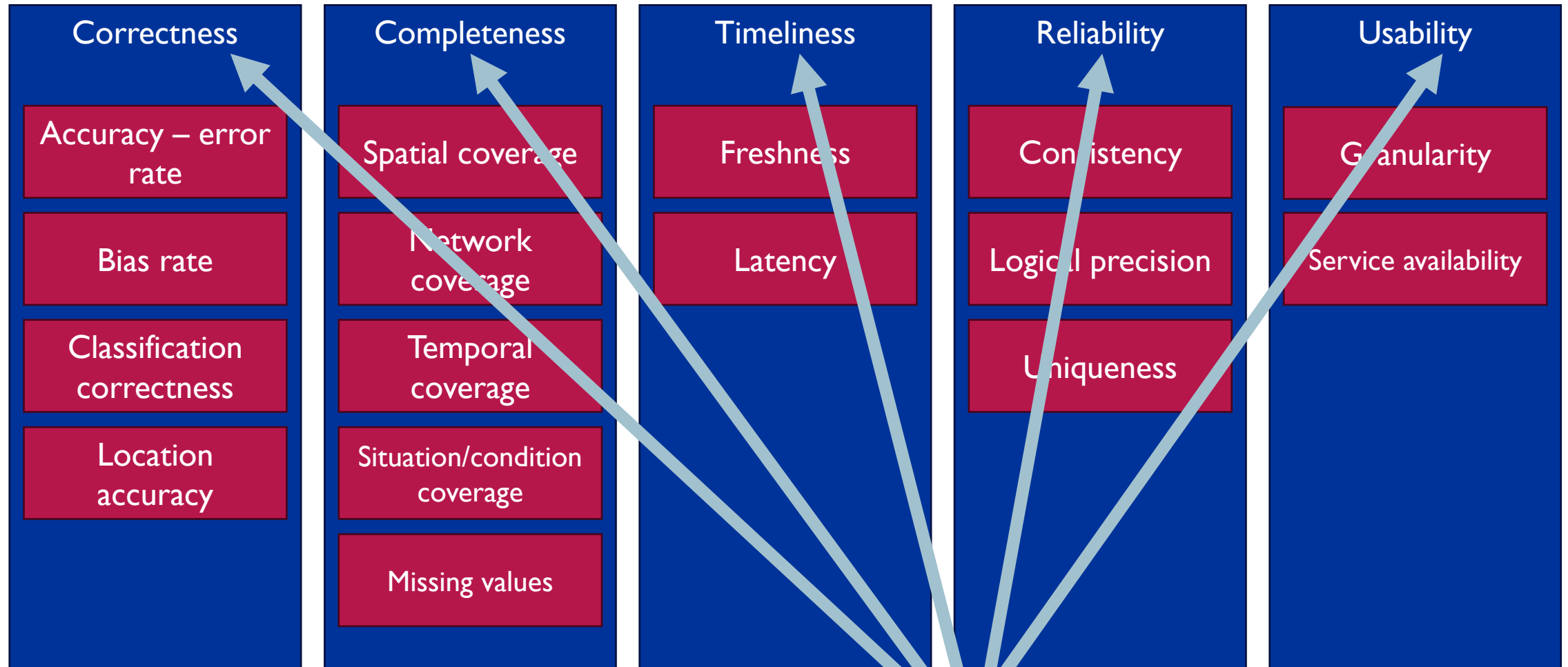
Techniques and procedures used to assess and measure the quality of data (i.e., quantify each criterion).

Data quality – Application area



<p>Process of observing a measurement sample by means of technical equipment (could also involve human activity).</p>	<p>Process of transferring the detected measurement sample from the measurement location to a central place.</p>	<p>The process of collecting several measurement samples from the measurement locations at a central entity.</p>	<p>Process of gathering the received measurement samples in a repository.</p>	<p>Process of combining raw data measurement samples from different means of detection into a representation of the traffic situation.</p>	<p>Process of checking on measurement samples and the reconstructed traffic situation with the goal of removing erroneous samples.</p>	<p>Process of transferring the content to a service provider, which will take care of the distribution of the content to the End Users.</p>
---	--	--	---	--	--	---

Data quality – Quality criteria (1/3)



Data quality – Quality criteria (2/3)

- **Accuracy – error rate:** extent to which data represents in an accurate & precise manner the real-world situation or conditions it describes without inaccuracies, discrepancies, or mistakes.
- **Bias rate:** extent to which data contains consistent inaccuracies, discrepancies, or mistakes compared to the true or expected values.
- **Classification correctness:** extent to which situations/events/conditions/vehicle types/... are classified correctly.
- **Location accuracy:** extent to which provided information is correctly georeferenced (i.e., provided locations accurately represent the real-world locations they intend to describe).
- **Spatial coverage:** spatial extent or range that a dataset or data resource encompasses.
- **Network coverage:** percentage of the applicable transportation network covered by a dataset or data resource.
- **Temporal coverage:** time period or duration for which a dataset or data source provides information or is relevant.
- **Situation/condition coverage:** variety of traffic conditions and situations captured and accurately represented within a data resource.

Data quality – Quality criteria (3/3)

- **Missing values:** number of entries that are absent or incomplete within a dataset or data source.
- **Freshness:** frequency with which data is updated.
- **Latency:** time delay or lag between the production (or update) of a data source and the moment this source is made available for use (through the NAP).
- **Consistency:** degree to which provided information is uniform and coherent throughout a dataset or database based on predefined rules or standards.
- **Logical precision:** extent to which a dataset or data source contains logical errors (e.g., traffic speeds greater than free flow speeds).
- **Uniqueness:** extent to which each data item or record is distinct and not duplicated within a dataset.
- **Granularity:** level of detail or resolution of the provided information based on the specific needs of the associated use cases.
- **Service availability:** extent to which a data service is operational and continuously accessible for users.

Data quality – Quality metrics (examples)

Dimension	Criterion	Metrics	Unit	Assessment Object
Correctness	Accuracy – error rate	$Q_{\text{Correlation Coefficient } R}(C)$ $Q_{\text{MAPE}}(C)$ $Q_{\text{RMPSE}}(C)$ $Q_{\text{RMPSE-weighted by length}}(C)$	-1 ... 1 0 ... ∞ 0 ... ∞ 0 ... ∞	Value pairs of reference data and test data
	Bias rate	$Q_{\text{MBPE}}(C)$	0 ... ∞	Value pairs of reference data and test data
Timeliness	Freshness	$Q_{\text{Update Frequency}}(C)$	Time interval	One data offer
	Latency	$Q_{\text{Latency of Availability}}(C)$	Time delta	One data offer
Completeness	Network coverage	$Q_{\text{Coverage ratio}}(D)$	0 ... 1	One data offer

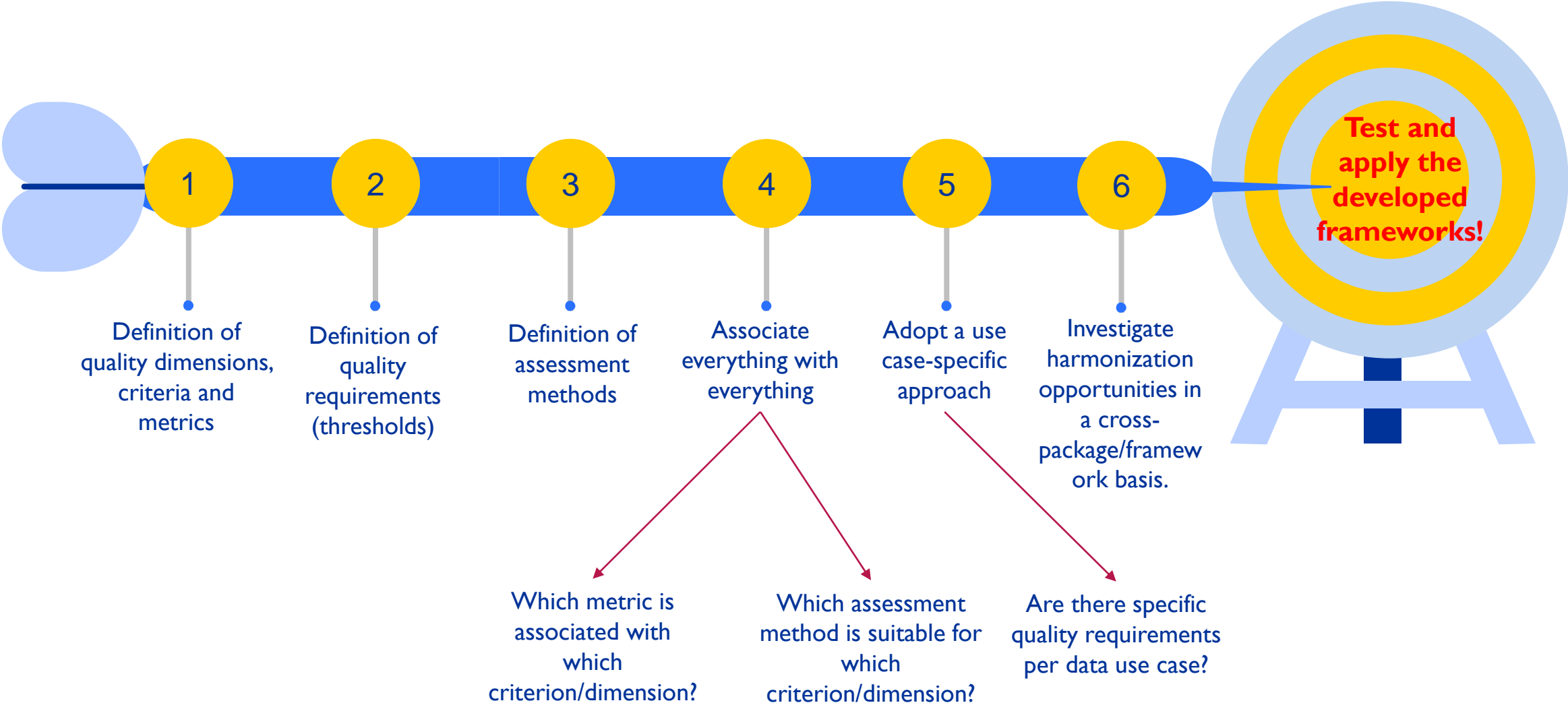
Data quality – Quality requirements (examples)

Dimension	Criterion	Quality level		
		Basic *	Enhanced **	Advanced ***
Accuracy	Precision – error rate	Correlation coefficient \geq 0.8 MAPE \leq 25%	Correlation coefficient \geq 0.9 MAPE \leq 10%	Correlation coefficient \geq 0.95 MAPE \leq 5%
	Bias rate
Timeliness	Freshness	\leq 15 min	\leq 5 min	\leq 1 min
	Latency	For 95 % of all reports: \leq 5 min	For 95 % of all reports: \leq 2 min	For 95 % of all reports: \leq 1 min
Completeness	Network coverage	Best effort	80%	90%

Data quality – Assessment methods

- **Continuous monitoring of equipment**
Goal: detect failures that are more or less specific for the type of equipment.
- **Manual verification of events or conditions**
Goal: verify manually reported events or conditions and take corrective actions (if needed).
- **Automated or regular monitoring of data correctness and latency**
Goal: evaluate processing performance and correctness in a continuous basis (via software solutions) or by drawing content samples in a regular basis & plan for improvements.
- **Reference testing**
Goal: make comparisons between an existing information service and a (typically purchased) ground-truth information service for a limited period of time in a time-space oriented context.
- **Surveys of perceived quality by users**
Goal: measure how the end users experience/perceive an information service. Data collection may be performed periodically (e.g., once a year).
- **Collection of direct user feedback**
Goal: actively interact with data users and collect feedback.
- **Monitoring of service use statistics**
Goal: collect indirect information by measuring the amount of service use through counters internet page visits, smartphone application downloads and use etc.
- **Others to be defined**

Data quality – Next steps



Thanks for your attention!

Contact information

Evangelos Mitsakis, emit@certh.gr

Chrysostomos Mylonas, chmylonas@certh.gr

Maria Stavara, mstavara@certh.gr